**Amendments to the Claims**:

This listing of claims will replace all prior versions, and listings, of claims in the application:

**Listing of Claims**:

1. (Currently amended) A method for generating a fingerprint of a document, performed by one or more server devices, the method comprising:

obtaining, by a processor associated with the one or more server devices, a plurality of overlapping blocks by sampling the document;

generating, by a processor associated with the one or more server devices, a set of checksum values [[for]] from the plurality of overlapping blocks;

choosing, by a processor associated with the one or more server devices, a subset of the set of checksum values, where the subset is less than an entirety of the set of checksum values;

initializing, by a processor associated with the one or more server devices, the fingerprint of the document by setting all bits of the fingerprint to zero;

addressing, by a processor associated with the one or more server devices, a particular bit of the fingerprint with a particular checksum value; and

flipping, by a processor associated with the one or more server devices, the particular bit of the fingerprint a number of times corresponding to a number of times the particular checksum value occurs in the subset

~~compacting the subset of the set of checksum values to obtain the fingerprint of the document by addressing bits in the fingerprint with particular values and flipping bits in the~~

~~fingerprint, where a number of times a particular bit is flipped is based on a number of checksum values in the subset that correspond to the particular value addressed to the particular bit~~.

2. (Canceled)

3. (Previously presented) The method of claim 1, where the fingerprint comprises a predetermined length.

4. (Previously presented) The method of claim 3, where the predetermined length is eight or sixteen bytes.

5. (Canceled)

6. (Currently amended) The method of claim [[5]] 1, where choosing a subset of the set of checksum values includes selecting a predetermined number of the smallest checksum values.

7. (Currently amended) The method of claim [[5]] 1, where choosing a subset of the set of checksum values includes selecting a predetermined number of the largest checksum values.

8. (Previously presented) The method of claim 1, further comprising:

hashing the subset of the set of checksum values to a length for addressing the bits of the fingerprint.

9. (Previously presented) The method of claim 8, where hashing the subset of the set of checksum values includes taking a number of least significant bits of the subset of the set of checksum values.

10. (Canceled)

11. (Previously presented) The method of claim 1, where each of the plurality of overlapping blocks is of a predetermined length.

12. (Previously presented) The method of claim 11, where obtaining a plurality of overlapping blocks further includes:

padding null characters to the document when a length of the document is below the predetermined length.

13. (Currently amended) A method for generating a fingerprint of a document, performed by one or more server devices, the method comprising:

sampling, by a processor associated with the one or more server devices, the document to obtain a plurality of overlapping samples;

generating, by a processor associated with the one or more server devices, a set of checksum values [[for]] from the plurality of overlapping samples;

selecting, by a processor associated with the one or more server devices, a ~~predetermined number~~ subset of the set of checksum values as those of the checksum values corresponding to a predetermined number of smallest checksum values or a predetermined number of largest checksum values;

addressing, by a processor associated with the one or more server devices, a particular bit of the fingerprint with a particular checksum value; and

~~setting bits in the fingerprint of the document by addressing bits in the fingerprint with particular values and~~ flipping, by a processor associated with the one or more server devices, ~~bits in the fingerprint, where a number of times a~~ the particular bit of the fingerprint a number of times ~~is flipped is based on a number of checksum values in the subset that correspond~~ corresponding to a number of times the particular checksum value ~~addressed to the particular bit~~ occurs in the subset.

14. (Previously presented) The method of claim 13, where the fingerprint comprises a predetermined length.

15. (Previously presented) The method of claim 14, where the predetermined length is eight or sixteen bytes.

16. (Canceled)

17. (Currently amended) The method of claim 13, further comprising:

hashing the ~~predetermined number~~ subset of the checksum values to a length for indexing the fingerprint of the document.

18. (Currently amended) The method of claim 17, where hashing the ~~predetermined number~~ subset of the checksum values includes taking a number of least significant bits of the predetermined number of checksum values.

19. (Currently amended) The method of claim 17, where ~~setting bits~~ flipping a particular bit in the fingerprint of the document includes flipping [[a]] the particular bit in the fingerprint of the document when the particular value addressed to the particular bit corresponds to a particular one of the hashed checksum values.

20. (Currently amended) A computer-implemented device comprising:

a memory to store instructions for implementing:

a fingerprint creation component to generate a fingerprint of a predetermined length for an input document, the fingerprint generated by

sampling the input document to obtain samples,

generating a set of checksum values [[for]] from the samples[[;]],

choosing a subset of the set of checksum values, where the subset is less than an entirety of the set of checksum values, [[and]]

~~generating the fingerprint from the subset of the set of checksum values by~~

addressing bits in the fingerprint with particular values, and

flipping ~~bits~~ a particular bit in the fingerprint~~, where~~ a number of times ~~a particular bit is flipped is~~ based on a number of checksum values in the subset that correspond to the particular value addressed to the particular bit, and

a similarity detection component to compare pairs of fingerprints to determine whether the pairs of fingerprints correspond to near-duplicate documents; and

a processor to execute the instructions in the memory.

21. (Previously presented) The computer-implemented device of claim 20, further including:

a search engine to return documents to a user as a single link when the documents are determined to correspond to near-duplicate documents.

22. (Previously presented) The computer-implemented device of claim 20, where the similarity detection component compares the pairs of fingerprints by calculating a hamming distance.

23. (Previously presented) The computer-implemented device of claim 22, where the similarity detection component determines that the pairs of documents correspond to near-duplicate documents when the hamming distance is below a threshold.

24. (Previously presented) The computer-implemented device of claim 20, where the fingerprint creation component additionally:

chooses the subset as a predetermined number of smallest checksums calculated from the subset of the samples.

25. (Previously presented) The computer-implemented device of claim 20, where the fingerprint creation component additionally:

chooses the subset as a predetermined number of largest checksums calculated from the subset of the samples.

26. (Currently amended) A computer-implemented device comprising:

memory to store instructions for implementing:

means for sampling a document to obtain a plurality of overlapping blocks,

means for generating a set of checksum values [[for]] from the plurality of overlapping blocks;

means for choosing a subset of the set of checksum values, where the subset is less than an entirety of the set of checksum values, and

means for ~~compacting the subset of the set of checksum values to obtain a a fingerprint of the document by~~ addressing bits in the fingerprint with particular values, and

means for flipping [[bits]] a particular bit in the fingerprint~~, where~~ a number of times ~~a particular bit is flipped is~~ based on a number of checksum values in the subset that correspond to the particular value addressed to the particular bit; and

a processor to execute the instructions in memory.

27. (Canceled)

28. (Previously presented) The computer-implemented device of claim 26, where the means for choosing a subset of the set of checksum values chooses the subset as a predetermined number of the smallest checksum values.

29. (Previously presented) The computer-implemented device of claim 26, where the means for choosing a subset of the set of checksum values chooses the subset as a predetermined number of the largest checksum values.

30. (Currently amended) The computer-implemented device of claim 26, ~~where the means for compacting the subset of the set of checksum values includes~~ further comprising:

means for hashing the checksum values.

31. (Currently amended) A ~~computer-readable~~ memory device containing program instructions that, when executed by a processor, cause the processor to:

sample a document to obtain a plurality of overlapping samples;

generate a set of checksum values [[for]] from the plurality of overlapping samples;

select a ~~predetermined number~~ subset of the set of checksum values as those of the checksum values corresponding to a predetermined number of smallest checksum values or a predetermined number of largest checksum values; [[and]]

~~set bits in a fingerprint of the document by~~ addressing bits in the fingerprint with particular values; and

flipping [[bits]] a particular bit in the fingerprint~~, where~~ a number of times ~~a particular bit is flipped is~~ based on a number of checksum values in the subset that correspond to the particular value addressed to the particular bit.

32. (Currently amended) The ~~computer-readable~~ memory device of claim 31, further including program instructions that, when executed by the processor, cause the processor to:

hash the ~~predetermined number~~ subset of the checksum values to a length for setting the bits in the fingerprint of the document.

33. (Currently amended) The ~~computer-readable~~ memory device of claim 32, where hashing the predetermined number of the checksum values includes taking a number of least significant bits of the predetermined number of checksum values.

34. (Currently amended) A computer-implemented device comprising:

memory to store instructions for implementing:

means for sampling a document to obtain a plurality of overlapping blocks,

means for calculating checksum values [[for]] from the plurality of overlapping blocks,

means for choosing a subset of the calculated checksum values, where the subset is less than an entirety of the calculated checksum values,

means for hashing the checksum values in the subset to a predetermined size;

means for addressing a particular bit of the fingerprint with a particular hashed checksum value; and

means for ~~setting bits in a c fingerprint of the document by~~ flipping [[bits]] the particular bit in the fingerprint~~, where a particular bit in the fingerprint is addressed with a particular hashed checksum value, and where~~ a number of times ~~the particular bit in the fingerprint is flipped corresponds~~ corresponding to a number of times the particular hashed checksum value occurs in the subset; and

a processor to execute the instructions in memory.

35. (Previously presented) The computer-implemented device of claim 34, further comprising:

means for initializing the fingerprint before setting bits in the fingerprint.

36. (Previously presented) The computer-implemented device of claim 34, further comprising:

means for comparing pairs of fingerprints to determine whether the pairs of fingerprints correspond to near-duplicate documents.